

**Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística**

Testes Estatísticos de Concordância

por

Maria Clara Vieira Borba

Relatório apresentado para
obtenção do título de bacharel em
Estatística.

Brasília

2015

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Testes Estatísticos de Concordância

por

Maria Clara Vieira Borba

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como parte dos requisitos para obtenção do título de

BACHAREL EM ESTATÍSTICA

Brasília, 2015.

Banca Examinadora:

Prof. Dr. Eduardo Yoshio Nakano - UnB (Orientador)

Prof. Dr. Cibele Queiroz da Silva - UnB

Prof. Dr. André Luiz Fernandes Cançado - UnB

Dedicatória

Aos meus pais, por me ensinarem que a melhor herança é a educação.

À minha querida avó, pelo amor incondicional.

Ao meu namorado, pelo carinho e incentivos diários.

Resumo

Este trabalho apresentou diversos procedimentos para verificar a concordância entre duas observações. Foram apresentados os principais testes de concordância disponíveis na literatura, classificando-os de acordo com o nível de mensuração da variável de interesse. Além dos testes tradicionais, este trabalho propôs procedimentos alternativos para a verificação da concordância através de métodos estatísticos cujo propósito original não é testar a concordância. Todos os testes apresentados foram ilustrados em exemplos envolvendo conjunto de dados gerados ou retirados da literatura e toda a análise foi realizada através do software R e os *scripts* utilizados para a realização dos testes foram apresentados no trabalho.

Palavras-chave: validação, teste-reteste, reprodutibilidade.

Abstract

This paper presented many procedures to verify the agreement between two observations. It presented the main agreement tests available in literature, sorting them according to the level of measure of the variable of interest. Besides traditional tests, this paper has proposed alternative procedures to verify the agreement through statistical methods whose original purpose are not test the agreement. All tests are shown illustrated in examples involving generated or taken from literature databases and the whole analysis was done through R software and the scripts used were presented in this paper.

Keywords: validation, test-retest, reproducibility.

Índice

1	Introdução	7
2	Revisão Bibliográfica	8
2.1	Coeficiente de Concordância de Kappa	8
2.2	Coeficiente de Correlação de Kendall	10
2.3	Coeficiente de Concordância de Kendall	13
2.4	Teste de Wilcoxon	14
2.5	Coeficiente de Correlação de Pearson	16
2.6	Teste t Pareado	18
2.7	Regressão Linear Simples	19
2.8	Gráficos de Bland-Altman	22
3	Ilustração dos Testes de Concordância em Dados Artificiais	24
3.1	Variáveis Nominais	24
3.1.1	Coeficiente de Concordância de Kappa	24
3.2	Variáveis Ordinais	26
3.2.1	Correlação de Kendall e teste de Wilcoxon	26
3.2.2	Coeficiente de Concordância de Kendall	29
3.3	Variáveis Quantitativas	31
3.3.1	Teste T-Pareado e Coeficiente de Correlação de Pearson	31
3.3.2	Regressão Linear Simples	34
4	Dois exemplos de aplicação dos testes de concordância	37
4.1	Aplicação 1 - Comparação de dois protótipos	37
4.2	Aplicação 2 - Concordância em notas atribuídas por dois professores . .	38
5	Conclusões	44
	Referências Bibliográficas	46

Capítulo 1

Introdução

Em muitas situações, deseja-se avaliar a concordância entre dois ou mais métodos que deveriam medir a mesma quantidade e isto pode ser feito através de testes estatísticos de concordância. Em pesquisas biomédicas, por exemplo, onde existem diferentes formas de avaliar uma determinada medida clínica, busca-se sempre novos métodos menos invasivos, com menores efeitos colaterais e que sejam mais simples ou mais baratos do que os métodos já consagrados.

Tem-se como objetivo, então, verificar se esses novos métodos são equivalentes a ponto de se poder substituir um pelo outro. Para que isso ocorra, os dois métodos em comparação devem apresentar viés próximo à zero e um erro que não tenha impacto significativo para que sejam permutáveis. Os testes de concordância também podem ser aplicados em estudos onde os níveis de confiabilidade teste-reteste são avaliados, em que uma mesma medida é realizada duas ou mais vezes com um intervalo entre as aplicações.

Existem diversos métodos de se avaliar a concordância entre medidas e esses variam de acordo com o nível de mensuração das variáveis. Desta forma, é necessário fazer uma discriminação das técnicas que pressupõem uma escala quantitativa (em geral, tem-se a suposição adicional de que as observações devem seguir uma distribuição Normal) com aquelas que podem ser utilizadas em uma escala qualitativa. Neste último caso, ainda é necessário subdividir as variáveis em dois critérios: ordinais e nominais.

Altman e Bland [16] mostraram em seu estudo casos em que técnicas como testes de comparação de médias, coeficientes de correlação e a técnica de regressão linear simples eram usadas de forma inadequada para a verificação da concordância. Neste contexto, serão propostos durante o trabalho procedimentos alternativos que permitirão a utilização das técnicas citadas anteriormente para a verificação de concordância.

Diante do exposto, o objetivo deste trabalho é apresentar diversos métodos de avaliação da concordância entre medidas, indicando em que situação cada um deles deve ser utilizado. Todos os testes apresentados serão ilustrados em exemplos fictícios e serão realizadas através do software R [3].

Capítulo 2

Revisão Bibliográfica

Este capítulo apresenta uma lista de técnicas estatísticas que podem ser utilizadas como testes de concordância. Entre outras, cita-se o coeficiente de concordância de Kappa, o coeficiente de correlação de Kendall, o coeficiente de concordância de Kendall, o teste de Wilcoxon, o coeficiente de correlação de Pearson, o teste T-pareado, a Regressão Linear Simples, e o gráfico de Bland-Altman [16].

Note que nem todos os testes apresentados são metodologias que visam verificar a concordância entre medidas (como por exemplo o teste T-Pareado, cujo objetivo é a comparação de duas médias de medidas pareadas). No entanto, a contextualização dessas metodologias dentro de um cenário de teste de concordância será apresentada nos capítulos seguintes deste trabalho.

2.1 Coeficiente de Concordância de Kappa

A estatística Kappa, baseada na precisão, fornece uma medida quantitativa da magnitude da concordância entre observadores, ou seja, o quanto as observações se afastam daquelas esperadas, frutos do acaso, indicando, assim, o quão legítimas as interpretações são [6].

Como uma melhor forma de ilustrar o desenvolvimento do processo de encontrar o coeficiente de concordância de Kappa, em um processo em que as respostas dos observadores são dicotômicas (por exemplo, sim ou não), tem-se que:

Tabela 2.1: Tabela de respostas dicotômicas para o coeficiente Kappa

		Resultado do Observador 1		
		Respostas	Sim Não	Total
Resultado do Observador 2	Sim	a	b	m_1
	Não	c	d	m_0
	Total	n_1	n_0	n

Fonte: Cohen, 1960.

onde,

- a e d representam o número de vezes que os observadores concordam.
- b e c representam o número de vezes que os observadores discordam.
- Se não houver discordância, b e c serão zero e a concordância observada p_o será 1.
- Se não houver concordância, a e d serão zero e a concordância observada p_o será 0.

Em que,

$$p_o = \frac{\text{número de concordâncias}}{\text{número de concordâncias} + \text{número de discordâncias}} \quad (2.1)$$

Para se encontrar a concordância esperada,

$$p_e = \left[\frac{n_1}{n} \times \frac{m_1}{n} \right] + \left[\frac{n_0}{n} \times \frac{m_0}{n} \right] \quad (2.2)$$

Pode-se, então, calcular o coeficiente Kappa (κ) por:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (2.3)$$

Os valores do coeficiente de Kappa variam entre 0 e 1, em que 0 representa não haver concordância além do puro acaso e 1 representa a concordância perfeita.

Então, para analisar a concordância entre duas medidas em relação ao coeficiente Kappa, de acordo com Landis e Koch [5], pode-se utilizar a seguinte interpretação:

Tabela 2.2: Tabela para a interpretação dos valores de Kappa

Valores obtidos de Kappa	Interpretação
$\kappa = 0,00$	Nenhuma Concordância
$0,00 < \kappa < 0,20$	Concordância Pobre
$0,20 \leq \kappa < 0,40$	Concordância Leve
$0,40 \leq \kappa < 0,60$	Concordância Moderada
$0,60 \leq \kappa < 0,80$	Concordância Substantial
$0,80 \leq \kappa \leq 1,00$	Concordância Quase Perfeita

Fonte: Landis e Koch, 1977.

O coeficiente κ é considerado uma medida mais robusta do que apenas o simples cálculo da porcentagem de concordância p_o , visto que o mesmo leva em consideração a concordância ao acaso.

No entanto, o valor de Kappa depende da predominância de uma variável em estudo. Uma grande prevalência resulta num alto nível de concordância esperada pelo acaso, o que resultará num valor κ mais baixo. Por sua vez, uma patologia de baixa prevalência, dará origem a valores de κ mais altos. Poderemos então cometer o erro de basear este índice na comparação de dois estudos com predominâncias distintas [7].

Testando as seguintes hipóteses:

H_0 : As respostas são discordantes ($\kappa = 0$).

H_a : As respostas são concordantes ($\kappa > 0$).

A significância do valor κ pode ser determinada através da estatística:

$$Z_\kappa = \frac{\kappa}{\sqrt{\frac{p_e}{n(1-p_e)}}} \approx N(0, 1). \quad (2.4)$$

O teste também pode ser realizado calculando-se o nível descritivo (p-valor):

$$p - \text{valor} = P(Z > Z_\kappa), \quad (2.5)$$

com Z tendo uma distribuição Normal Padrão. Nesse caso, fixando um nível de significância α , a hipótese de concordância das respostas (H_0) é rejeitada se $p - \text{valor} < \alpha$.

2.2 Coeficiente de Correlação de Kendall

De acordo com SalKind e Neil [9], o coeficiente de correlação de Kendall por postos avalia o grau de similaridade entre dois grupos (X e Y) de *rankings* dados pelos mesmos

grupos de variáveis, que precisam, no mínimo, ser ordinais. Esse coeficiente depende do número de “inversões” de pares de objetos que podem ser necessários para transformar um *ranking* em outro, mas deve-se levar em consideração que não há suposição de normalidade dos dados.

Para se fazer isso, cada ordem de postos é representada por uma combinação de todos os pares e os valores 0 e 1 definem o par quando sua ordem corresponde ou não com o modo que as duas variáveis foram ordenadas.

Quando a ordem dos elementos do conjunto é levada em consideração, obtém-se um conjunto ordenado que também pode ser representado pelo *ranking* dos seus elementos.

Um conjunto ordenado de n elementos pode ser decomposto em $\frac{1}{2}n(n-1)$ possíveis pares ordenados.

Para comparar dois conjuntos ordenados (no mesmo conjunto de elementos), a abordagem de Kendall é contar o número de pares diferentes entre dois conjuntos ordenados. Esse número representa a distância entre esses conjuntos, chamada diferença simétrica da distância (*symmetric difference distance*). A diferença simétrica é um conjunto de operações que associam para dois conjuntos, os conjuntos de elementos que pertencem a apenas um conjunto.

O coeficiente de correlação de Kendall é obtido normalizando a diferença simétrica, assim, apresentará valores entre -1 e $+1$, com -1 representando a maior distância possível e $+1$ representando a menor distância possível.

Para o cálculo do coeficiente de correlação de Kendall, definimos as quantidades N_C e N_E que representam, respectivamente, o número de pares com a ordenação correta e o número de pares com a ordenação errada. Levando em consideração que o número máximo de pares ordenados existentes é $\frac{1}{2}n(n-1)$, dá-se a seguinte fórmula para a correlação de Kendall, sem empates:

$$\tau_{Kendall} = \frac{N_C - N_E}{\frac{1}{2}n(n-1)}, \quad (2.6)$$

E, quando há empates,

$$\tau_{Kendall} = \frac{N_C - N_E}{\frac{1}{2}\sqrt{n(n-1) - \sum_{i=1}^I t_{x;i}(t_{x;i} - 1)}\sqrt{n(n-1) - \sum_{j=1}^J t_{y;j}(t_{y;j} - 1)}} \quad (2.7)$$

em que N_C e N_E são, respectivamente, o número de pares com a ordenação correta e o número de pares com a ordenação errada, $t_{x;i}$ é o número de elementos do i -ésimo empate da variável X e $t_{y;j}$ é o número de elementos do j -ésimo empate da variável Y .

Note que:

- Se todos os pares estiverem na ordenação correta, $\tau_{Kendall} = 1$.
- Se todos os pares estiverem na ordenação errada, $\tau_{Kendall} = -1$.
- Se metade dos valores estiverem na ordenação correta e metade na errada, então não há nenhuma associação entre as variáveis e $\tau_{Kendall} = 0$.

Então, considerando as seguintes hipóteses,

H_0 : Não existe associação entre as duas variáveis ($\tau_{Kendall} = 0$).

H_a : Existe associação entre as duas variáveis ($\tau_{Kendall} \neq 0$).

a estatística do teste é

$$Z_{Kendall} = \frac{N_C - N_E}{\sqrt{\frac{v_0 - v_x - v_y}{18} + \frac{v_1}{2n(n-1)} + \frac{v_2}{9n(n-1)(n-2)}}} \approx N(0, 1) \quad (2.8)$$

,

em que

$$v_0 = n(n-1)(2n+5)$$

$$v_x = \sum_{i=1}^I t_{x;i}(t_{x;i} - 1)(2t_{x;i} + 5)$$

$$v_y = \sum_{j=1}^J t_{y;j}(t_{y;j} - 1)(2t_{y;j} + 5)$$

$$v_1 = \left(\sum_{i=1}^I t_{x;i}(t_{x;i} - 1)\right)\left(\sum_{j=1}^J t_{y;j}(t_{y;j} - 1)\right)$$

$$v_2 = \left(\sum_{i=1}^I t_{x;i}(t_{x;i} - 1)(t_{x;i} - 2)\right)\left(\sum_{j=1}^J t_{y;j}(t_{y;j} - 1)\sum_{j=1}^J t_{y;j}(t_{y;j} - 1)(t_{y;j} - 2)\right)$$

$t_{x;i}$ é o número de elementos do i -ésimo empate da variável X e

$t_{y;j}$ é o número de elementos do j -ésimo empate da variável Y .

Para se fazer a decisão sobre a aceitação ou rejeição de determinada hipótese, utiliza-se o p-valor:

$$p - \text{valor} = 2 \times P(Z > |Z_{Kendall}|) \quad (2.9)$$

com Z tendo uma distribuição Normal padrão.

Fixando um nível de significância α , rejeita-se a hipótese $H_0 : \tau_{Kendall} = 0$ se $p - valor < \alpha$.

Kendall [9] sugere que um tamanho de amostra $n \geq 8$ já apresenta uma boa aproximação para a distribuição Normal.

2.3 Coeficiente de Concordância de Kendall

O coeficiente de concordância de Kendall ($W_{Kendall}$) é uma medida de associação entre k conjuntos de postos de N objetos ou indivíduos, que assume valores entre 0 e 1, sendo 1 a representação da concordância perfeita.

O coeficiente $W_{Kendall}$ não pode tomar valores negativos pelo fato de considerar mais de dois conjuntos de postos e, quando isso acontece, esses conjuntos não podem ser completamente discordantes. Por exemplo, se X e Y discordam, e se X também discorda de Z , então Y e Z certamente concordarão, já que quando se considera mais de dois conjuntos, concordância e discordância não são opostos [4].

Para calcular o coeficiente de concordância de Kendall, determina-se a soma dos postos, R_j , em cada coluna de uma tábua $k \times N$.

Tabela 2.3: Tabela representativa para os valores da concordância de Kendall

Postos	Indivíduos			
	1	2	...	N
1	.	.		.
2	.	.		.
...				
k	.	.		.
R_j	R_1	R_2	...	R_k

Em seguida, soma-se os R_j 's e os divide por N , obtendo a média dos R_j . Cada um destes pode, então, ser expresso como um desvio em relação à média, o que leva à interpretação da relação entre os postos, já que quanto maior são os desvios, maior é o grau de associação entre os k conjuntos de postos. Por fim, determina-se s , a soma dos quadrados desses desvios. Assim, no caso sem empates, o coeficiente de concordância de Kendall é dado por

$$W_{Kendall} = \frac{s}{\frac{1}{12}k^2(N^3 - N)}, \quad (2.10)$$

e, quando há empates,

$$W_{Kendall} = \frac{s}{\frac{1}{12}k^2(N^3 - N) - \frac{k}{12} \sum_{j=1}^k \sum_{i=1}^I (t_{j;i}^3 - t_{j;i})} \quad (2.11)$$

em que,

$$s = \sum_{j=1}^k \left(R_j - \frac{\sum_{j=1}^k R_j}{N} \right)^2;$$

k : número de conjuntos de postos;

N : número de objetos ou indivíduos a que se atribuíram postos;

R_j : soma dos postos da coluna j , $j = 1, 2, \dots, k$

$t_{j;i}$ é o número de elementos do i -ésimo empate da coluna j .

Para testar a significância do coeficiente de concordância de Kendall, consideramos as seguintes hipóteses:

H_0 : Não há concordância entre os k grupos ($W_{Kendall} = 0$).

H_a : Há concordância entre os k grupos ($W_{Kendall} \neq 0$).

A significância pode ser determinada através da estatística

$$Q_{Kendall}^2 = k(N-1)W_{Kendall} \approx \chi_{(N-1)}^2 \quad (2.12)$$

O nível descritivo do teste (p -valor) é dado por

$$p - valor = P(\chi_{(N-1)}^2 > Q_{Kendall}^2) \quad (2.13)$$

que é obtido através de uma distribuição qui-quadrado com $(N-1)$ graus de liberdade. Neste caso, fixando um nível de significância α , a hipótese de discordância ($H_0 : W_{Kendall} = 0$) é rejeitada se $p - valor < \alpha$.

2.4 Teste de Wilcoxon

O teste de Wilcoxon tem como objetivo comparar dois grupos (X e Y) pareados, cujos dados devem ser, no mínimo, ordinais. A lógica do teste de Wilcoxon é bem simples. Os dados são classificados de forma a produzir dois grupos, um para cada condição a que se deseja estudar. Se existir uma diferença sistemática entre as duas condições, o posto mais alto pertencerá à uma condição e o mais baixo dos postos pertencerá à outra condição.

Para o cálculo da estatística de Wilcoxon, primeiramente calcula-se a diferença dos valores (ou postos) das duas variáveis. Eventualmente os escores dos pares podem ser

iguais e, portanto, resultar em uma diferença nula. Neste caso esse par é excluído da análise. A seguir, obtém-se os postos do valor absoluto das diferenças não nulas. Pode-se, assim, calcular S_p e S_n , que representam a soma dos postos das diferenças positivas e negativas, respectivamente.

Sob a hipótese de não existir diferença entre as variáveis, ou seja,

H_0 : Os grupos X e Y são iguais

H_a : Os grupos X e Y são diferentes.

Além dos empates dos escores dos pares, pode ocorrer um outro tipo de empate. Duas ou mais diferenças (não nulas) podem ter o mesmo valor absoluto. Neste caso, atribui-se o mesmo posto a esses empates.

Na situação em que não há empates, a estatística do teste de Wilcoxon, $W_{Wilcoxon}$, é simplesmente o menor valor entre S_p e S_n , isto é,

$$W = \min\{S_p, S_n\}. \quad (2.14)$$

Quanto menor o valor de $W_{Wilcoxon}$, menor é a probabilidade dos postos das diferenças terem ocorrido ao acaso. Assim, a hipótese de não existir diferenças (H_0) é rejeitada se $W < W_c$, onde W_c é o valor crítico encontrado na tabela do teste de Wilcoxon.

Tabela 2.4: Valores críticos de W

N	Nível de signif. teste Unilateral		
	0,025	0,01	0,005
	Nível de signif. teste Bilateral		
	0,05	0,02	001
6	0	-	-
7	2	0	-
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

No caso de empates, se a amostra for grande o suficiente, o teste de Wilcoxon pode ser aproximado pela distribuição normal. Nesta situação, a estatística do teste é alterada pela forma:

$$Z_{Wilcoxon} = \frac{\min(S_p, S_n) - \frac{m(m+1)}{4}}{\sqrt{\frac{m(m+1)(2m+1)}{24} - \frac{\sum_{j=1}^l (t_j^3 - t_j)}{48}}} \approx N(0, 1) \quad (2.15)$$

em que,

S_p : soma dos postos das diferenças positivas.

S_n : soma dos postos das diferenças negativas.

m : número de diferenças diferentes de zero.

l : número de empates (empates dos postos).

t_j : número de elementos no j -ésimo empate.

Assim, o p -valor pode ser calculado da seguinte forma:

$$p - \text{valor} = 2 \times P(Z > |Z_{Wilcoxon}|) \quad (2.16)$$

com Z tendo uma distribuição normal padrão. Logo, se $p - \text{valor} < \alpha$, rejeita-se a hipótese de igualdade dos grupos (H_0), onde α é o nível de significância do teste.

2.5 Coeficiente de Correlação de Pearson

Para Moore [14] “a correlação mensura a direção e o grau da relação linear entre duas variáveis quantitativas”.

Duas variáveis se associam quando elas guardam semelhanças na distribuição de seus escores. No caso da correlação de Pearson, ele é uma medida da variância compartilhada entre duas variáveis. Por outro lado, modelo linear supõe que o aumento ou decremento de uma unidade na variável X gera o mesmo impacto em Y . Em termos gráficos, por relação linear, entende-se que a melhor forma de ilustrar o padrão de relacionamento entre duas variáveis é através de uma linha reta. Portanto, a correlação de Pearson exige um compartilhamento de variância e que essa variação seja distribuída linearmente.

Assim, o coeficiente de correlação de Pearson, $r_{Pearson}$, é uma medida de associação linear entre as duas variáveis X e Y e apresenta a seguinte fórmula:

$$r_{Pearson} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.17)$$

com $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ e n é o tamanho da amostra.

O coeficiente de correlação de Pearson varia de -1 a 1. O sinal indica direção positiva ou negativa do relacionamento e o valor sugere a força da relação entre as variáveis. Uma correlação perfeita (-1 ou 1) indica que o escore de uma variável pode ser determinado exatamente ao se saber o escore da outra. Por outro lado, uma correlação de valor zero indica que não há relação linear entre as variáveis.

A interpretação da magnitude dos coeficientes varia para cada pesquisador. Em geral, quanto mais perto de 1 (independentemente do sinal), maior é o grau de dependência estatística linear entre as variáveis e quanto mais perto de zero, menor é a força dessa relação.

Algumas condições precisam ser satisfeitas para que o uso dessa medida seja apropriado:

- As variáveis devem ser quantitativas.
- Os valores observados precisam estar normalmente distribuídos. Esse pressuposto é importante principalmente em amostras pequenas ($N < 40$), já que, a partir do Teorema do Limite Central, sabe-se que na medida em que o número de observações aumenta, a distribuição das médias amostrais se aproxima da curva normal, independente do formato da distribuição dos dados na população.
- É necessário uma análise de outliers, pelo fato de o coeficiente de correlação ser fortemente afetado pela presença deles.
- Os pares de observações devem ser independentes entre si.

O coeficiente $r_{Pearson}$ é fortemente influenciado pela média da distribuição. Por esse motivo, um dos pressupostos centrais para que essa medida seja adequadamente utilizada é de que as observações obedeçam a uma distribuição normal.

Assim, sob as hipóteses

H_0 : Não existe associação linear entre as variáveis ($r_{Pearson} = 0$)

H_a : Existe associação linear entre as variáveis ($r_{Pearson} \neq 0$)

A estatística do teste é dada por:

$$t_{Pearson} = \frac{r_{Pearson} \sqrt{n-2}}{\sqrt{1-r_{Pearson}^2}} \approx t_{(n-2)} \quad (2.18)$$

A decisão de rejeitar a hipótese nula pode ser tomada por meio do nível descritivo (p -valor), que pode ser calculado por:

$$p - \text{valor} = 2 \times P(T_{(n-2)} > |t_{\text{Pearson}}|) \quad (2.19)$$

com $T_{(n-2)}$ tendo uma distribuição t de Student com $(n-2)$ graus de liberdade. Assim, considerando um nível de significância α , a hipótese nula é rejeitada se $p - \text{valor} < \alpha$.

2.6 Teste t Pareado

O teste t pareado é útil quando se deseja comparar duas médias em que, para cada unidade amostral, é realizada duas medições (X e Y) da característica de interesse. Considere duas amostras independentes X_1, \dots, X_n e Y_1, \dots, Y_n . Se as observações são pareadas, pode-se considerar que existe, na realidade, uma amostra de pares $(X_1, Y_1), \dots, (X_n, Y_n)$. Definindo $D_i = X_i - Y_i$, para $i = 1, \dots, n$, tem-se a amostra D_1, \dots, D_n , resultante das diferenças entre os valores de cada par. Para a realização do teste t pareado, supõe-se que as diferenças D_i são provenientes de uma distribuição Normal com média μ_D e variância σ_D^2 , isto é, $D_i \sim N(\mu_D, \sigma_D^2)$.

O parâmetro μ_D é estimado pela média amostral das diferenças ($\bar{D} = \frac{\sum_{i=1}^n D_i}{n}$) e o parâmetro σ_D^2 é estimado pela variância amostral das diferenças, ou seja,

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n - 1}$$

Para realizar o teste de diferença das médias, tem-se a seguinte hipótese bilateral:

H_0 : As médias de X e Y são iguais ($\mu_D = 0$)

H_a : As médias de X e Y são diferentes ($\mu_D \neq 0$)

A estatística do teste é dada por:

$$T_{\text{obs}} = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \sim t_{(n-1)} \quad (2.20)$$

A decisão de rejeitar a hipótese nula pode ser tomada por meio do nível descritivo (p -valor), que pode ser calculado por:

$$p - \text{valor} = 2P(T_{(n-2)} > |T_{\text{obs}}|) \quad (2.21)$$

com $T_{(n-1)}$ tendo uma distribuição t de Student com $(n-1)$ graus de liberdade. Assim, considerando um nível de significância α , a hipótese nula é rejeitada se $p - \text{valor} < \alpha$.

2.7 Regressão Linear Simples

A regressão linear é uma metodologia estatística que utiliza a relação entre duas ou mais variáveis, tanto qualitativas quanto quantitativas, de tal forma que uma variável pode ser predita a partir das outras.

Regressão é o estudo da dependência. A análise de regressão responde perguntas sobre a dependência de variáveis respostas em uma ou várias estimativas, incluindo a predição de valores futuros de uma resposta, descobrindo quais estimativas são importantes, e estimando o impacto de mudar a estimativa ou o tratamento no valor da resposta.

O caso mais simples de regressão é quando se tem apenas duas variáveis e a relação entre elas pode ser representada por uma linha reta, a chamada Regressão Linear Simples.

Tem-se como objetivo, então, avaliar uma possível dependência de uma variável resposta em relação à outra variável e expressar esta relação por meio de uma equação linear.

A equação geral de acordo com Weisberg [15], é a seguinte:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (2.22)$$

com $i = 1, \dots, n$,

- Y_i é a variável resposta ou dependente.
- X_i é a variável independente.
- β_0 é o coeficiente linear ou intercepto.
- β_1 é o coeficiente angular ou parâmetro de regressão.
- ϵ_i é o erro amostral, independentes e com distribuição $N(0, \sigma^2)$

Deve-se supor, em uma regressão linear, que os erros aleatórios apresentem em média zero, isto é, $E(\epsilon_i) = 0$.

Assumindo que a $\mathbb{E}(\epsilon_i) = 0$, a função resposta para o modelo de regressão 2.22 é:

$$\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i \quad \text{e} \quad \text{Var}(Y_i) = \sigma^2$$

Estimando-se, então, os parâmetros β_0 e β_1 pelo Método dos Mínimos Quadrados:

Utilizando as definições de Y_i e da $\mathbb{E}(Y_i)$, pode-se dizer que

$$\epsilon_i = Y_i - \mathbb{E}(Y_i)$$

Assim,

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2$$

Calculando as derivadas em relação à β_0 e à β_1 e igualando-as à zero,

$$\frac{\partial}{\partial \beta_0} \left[\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \right] = -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\frac{\partial}{\partial \beta_1} \left[\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \right] = 2 \sum_{i=1}^n (-X_i)(Y_i - \beta_0 - \beta_1 X_i) = 0$$

Resolvendo o seguinte sistema:

$$\begin{cases} \sum_{i=1}^n Y_i = n\beta_0 + \beta_1 \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i Y_i = \beta_0 \sum_{i=1}^n X_i + \beta_1 \sum_{i=1}^n X_i^2 \end{cases}$$

Encontra-se, assim, os estimadores dos parâmetros β_0 e β_1 :

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sum_{i=1}^n X_i^2 - n \bar{X}^2}$$

Com $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ e $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$.

O parâmetro β_1 indica a mudança na resposta média $\mathbb{E}(Y)$ com o aumento de uma unidade da variável X .

Regressão sem intercepto

Agora, durante o trabalho será necessário entender o modelo de regressão linear simples em que $\beta_0 = 0$, ou seja, um modelo sem intercepto. Desta forma, o estimador pode ser encontrado da mesma maneira de um modelo com intercepto, pelo método dos Mínimos Quadrados.

$$\epsilon_i = Y_i - \beta_1 X_i$$

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_1 X_i)^2$$

Agora, derivando-se em relação apenas ao β_1 e igualando à zero,

$$\frac{d}{d\beta_1} \left[\sum_{i=1}^n (Y_i - \beta_1 X_i)^2 \right] = 0$$

$$2 \sum_{i=1}^n (-X_i)(Y_i - \beta_1 X_i) = 0$$

Logo,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$$

Pode-se considerar que não há uma dependência da variável resposta se o valor de $\hat{\beta}_1$ for igual à zero.

A definição de um teste de hipóteses, $H_0 : \beta_1 = 1$, para um modelo de regressão linear simples sem intercepto é de grande importância, para que ele possa ser utilizado para analisar a concordância de medidas entre duas variáveis.

Considerando um modelo de regressão simples sem intercepto e a seguinte hipótese

$$H_0: \beta_1 = 1$$

$$H_a: \beta_1 \neq 1$$

tem-se que a estatística do teste é dada por:

$$T = \frac{\hat{\beta}_1 - 1}{\sqrt{V(\hat{\beta}_1)}} \sim t_{(n-2)} \quad (2.23)$$

em que $\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$ é a estimativa do coeficiente angular da regressão sem intercepto,

cujas variância é estimada por $V(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{n(\sum_{i=1}^n X_i^2 - n\bar{X}^2)}$. Aqui, $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{\beta}_1 X_i)^2}{n-2}$.

A decisão de rejeitar a hipótese nula pode ser tomada por meio do nível descritivo (p -valor), que pode ser calculado por:

$$p - \text{valor} = 2P(T_{(n-2)} > |T_{obs}|) \quad (2.24)$$

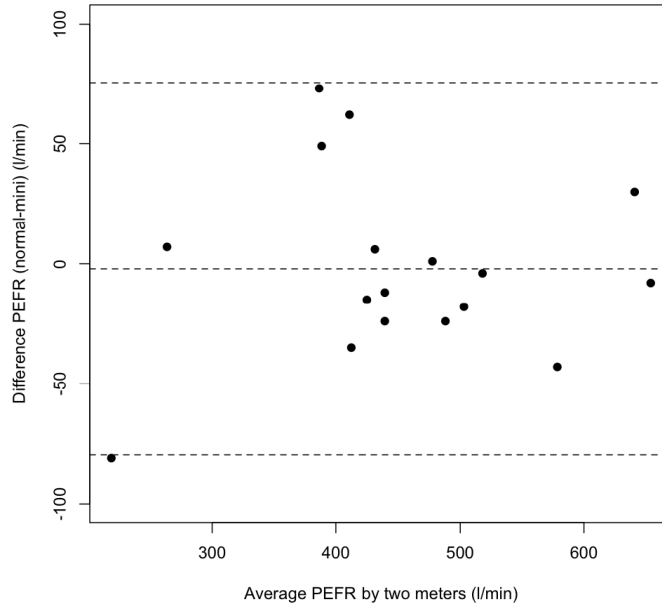
Com $T_{(n-2)}$ tendo uma distribuição t de Student com $(n - 2)$ graus de liberdade. Assim, considerando um nível de significância α , a hipótese nula é rejeitada se $p - \text{valor} < \alpha$.

2.8 Gráficos de Bland-Altman

Os gráficos de Bland-Altman [16] analisam a concordância entre duas variáveis quantitativas, X e Y , a partir de uma visualização gráfica através de um gráfico de dispersão entre a diferença de duas variáveis $(X - Y)$ e a média das duas $(X + Y)/2$, onde é possível visualizar o viés e o erro, além de *outliers* e tendências.

Sabendo que as diferenças entre métodos são independentes, esses podem ser comparados de forma simples, analisando a diferença de um pelo outro individualmente. A média dessas diferenças serão vieses relativos e seus desvios padrões será a estimativa do erro.

Utiliza-se um gráfico de diferenças entre as variáveis contra as médias das mesmas, como a figura 2.1, nos permite também investigar uma possível relação entre o erro e o verdadeiro valor. Não se sabe o valor verdadeiro, assim, a média das duas medidas é a melhor estimativa que se tem.



Fonte: J. Martin Bland e Douglas G. Altman [17].

Figura 2.1: Exemplo de Gráfico de Bland-Altman

Considerando a figura (2.1), a um nível de significância de $(1 - \alpha) \times 100\%$, a linha pontilhada central representa a média das diferenças entre duas variáveis e é definida por:

$$LC = \bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$$

com $D_i = (X_i - Y_i)$.

As duas linhas pontilhadas mais extremas do gráfico são os limites superior e inferior de Bland-Altman, que são definidos por:

$$LS = \bar{D} + t_{(n-1; 1-\frac{\alpha}{2})} \frac{S_D}{\sqrt{n}}$$

$$LI = \bar{D} - t_{(n-1; 1-\frac{\alpha}{2})} \frac{S_D}{\sqrt{n}}$$

em que $D_i = (X_i - Y_i)$, $S_D = \frac{\sqrt{\sum_{i=1}^n (D_i - \bar{D})^2}}{n-1}$ e $t_{(n-1; 1-\frac{\alpha}{2})}$ é o quantil $(1 - \frac{\alpha}{2})$ de uma distribuição t-Student com $n - 1$ graus de liberdade.

Se há uma variação considerável de resultados ao se repetir um método, a concordância entre ele e outro método tende a ser pobre [17].

A diferença $(X - Y)$ estar dentro dos limites do intervalo significa concordância entre X e Y .

Capítulo 3

Ilustração dos Testes de Concordância em Dados Artificiais

Tendo como base a Revisão Bibliográfica realizada e disposta no Capítulo 2, pode-se considerar que todos os testes citados apresentam algum tipo de restrição, principalmente quanto ao nível de mensuração da variável analisada. Assim, é necessário discriminar as técnicas que pressupõem uma escala quantitativa (em geral, tem-se a suposição adicional de que as observações devam seguir uma distribuição Normal) com aquelas que podem ser utilizadas em uma escala qualitativa. Neste último caso, ainda é necessário subdividir as variáveis em dois critérios: ordinais e nominais.

Neste contexto, este capítulo apresentará uma ilustração da aplicação dos testes de concordância, classificando-os de acordo com o nível de mensuração da variável considerada. Os testes serão aplicados em conjuntos de dados fictícios, gerados exclusivamente para ilustrar os procedimentos dos testes. Os *scripts* para a execução dos testes no R são apresentados após a ilustração de cada teste.

3.1 Variáveis Nominais

3.1.1 Coeficiente de Concordância de Kappa

O coeficiente de concordância de Kappa é utilizado para analisar a concordância entre duas variáveis nominais dicotômicas. A concordância pode ser verificada pelo coeficiente de Kappa quando $\kappa \approx 1$, cuja significância pode ser verificada, ao rejeitar a hipótese nula de que as hipóteses são discordantes ($H_0 : \kappa = 0$).

Considere um exemplo em que se deseja verificar a reprodutibilidade (concordância) de uma questão cuja resposta é dicotômica (sim ou não). Suponha que a mesma pergunta foi feita a cada respondente em dois momentos distintos e separados por um determinado período de tempo. As respostas obtidas são apresentadas na tabela abaixo.

Tabela 3.1: Tabela de respostas dicotômicas para o coeficiente Kappa

		Depois		
Respostas		Sim	Não	Total
Antes	Sim	49	17	66
	Não	23	61	84
	Total	72	78	150

Objetivo é saber se as respostas são concordantes, isto é, testar a hipótese $H_a : \kappa > 0$.

Iniciando a análise, deve-se encontrar a concordância observada, p_o , e esperada, p_e , cujas expressões são dadas, respectivamente, por (2.1) e (2.2).

Assim,

$$p_o = \frac{49 + 61}{150} = 0,7333$$

$$p_e = \left[\frac{72}{150} \times \frac{66}{150} \right] + \left[\frac{78}{150} \times \frac{84}{150} \right] = 0,5024$$

Logo, a estatística do teste, Kappa, pode ser calculada através de (2.3):

$$\kappa = \frac{0,7333 - 0,5024}{1 - 0,5024} = 0,4640$$

Apenas com o valor de κ e utilizando a tabela 2.2 como referência, pode-se afirmar que as respostas apresentam uma concordância moderada entre si.

A estatística Z_κ , descrita por (2.4), pode determinar a significância do valor do coeficiente de Kappa:

$$Z_\kappa = \frac{0,4640}{\sqrt{\frac{0,5024}{150(1-0,5024)}}} \approx N(0,1) \quad (3.1)$$

E, conseqüentemente, o p-valor pode ser encontrado,

$$p - \text{valor} = P(Z > Z_\kappa) = P(Z > 5,6556) \approx 0 \quad (3.2)$$

com Z tendo uma distribuição Normal Padrão. Nesse caso, fixando um nível de significância $\alpha = 0,05$, a hipótese de concordância das respostas (H_0) é rejeitada, já que $p - \text{valor} = 0 < 0,05 = \alpha$.

Script do R

```
dados <- matrix(c(49,23,17,61),nrow=2)
```

```
install.packages("psych")  
require(psych)
```

```
k <- wkappa(dados)
```

3.2 Variáveis Ordinais

3.2.1 Correlação de Kendall e teste de Wilcoxon

Os testes de correlação de Kendall e Wilcoxon podem ser utilizados para verificar a associação entre variáveis ordinais. Cada um desses testes não possibilita a inferência sobre a concordância das variáveis analisadas, quando usados separadamente.

O teste de Kendall tem a capacidade apenas de provar a discordância entre variáveis. Isso acontece quando $\tau_{Kendall} \approx 0$ ou $\tau_{Kendall} < 0$, ou seja, quando se aceita a hipótese de que as variáveis não estão associadas ou apresentam associação negativa. Primeiramente, se metade dos valores estiverem na ordenação correta e metade na errada, então não há nenhuma associação entre as variáveis, o que leva também à discordância. De outra forma, uma associação negativa implica diretamente em uma discordância.

Assim, utilizar o coeficiente de correlação de Kendall em duas variáveis ordinais permite apenas verificar a falta de concordância entre elas. Um coeficiente de correlação próximo a 1 indica que existe uma associação entre as duas variáveis, mas essa associação, não necessariamente, implica concordância.

Independentemente do resultado do teste de Wilcoxon, quando o teste de Kendall não rejeita a hipótese nula de discordância, pode-se afirmar que as variáveis são discordantes.

Como exemplo, considere os conjuntos de valores discordantes apresentados pela Tabela 3.2.

Utilizando os dados da tabela 3.2, item (a), calcula-se o coeficiente de correlação Kendall, que resultou em $\tau_{Kendall} = -1$, indicando uma associação negativa entre X e Y . Assim, as variáveis são discordantes, além do teste de Wilcoxon ter resultado $W = 0$ indicando que existe diferença entre as duas variáveis.

Utilizando ainda a mesma tabela, agora item (b), encontra-se $\tau_{Kendall} = -0,111$, confirmando que as variáveis são discordantes, mesmo quando o resultado de Wilcoxon é $W = 50$, o qual não rejeita a hipótese nula de que não há diferenças entre as variáveis. Logo, com o teste de correlação de Kendall é possível verificar a discordância entre duas observações.

Da mesma forma, independentemente do resultado do coeficiente de correlação de Kendall, quando o teste de Wilcoxon rejeita a hipótese de que não há diferenças entre as

Tabela 3.2: Valores das variáveis X e Y para o teste de Kendall

(a) Wilcoxon rejeita H_0		(b) Wilcoxon não rejeita H_0	
X	Y	X	Y
1	99	1	10
2	98	2	1
3	97	3	9
4	96	4	2
5	95	5	8
6	94	6	3
7	93	7	7
8	92	8	4
9	91	9	6
10	90	10	5

variáveis, o mesmo consegue comprovar a discordância entre as duas medidas ordinais. A Tabela 3.3 apresenta conjuntos de dados discordantes que ilustram essa situação.

Tabela 3.3: Valores das variáveis X e Y para o teste de Wilcoxon

(a) Kendall não rejeita H_0		(b) Kendall rejeita H_0	
X	Y	X	Y
1	99	1	11
2	98	2	12
3	97	3	13
4	96	4	14
5	95	5	15
6	94	6	16
7	93	7	17
8	92	8	18
9	91	9	19
10	90	10	20

Utilizando os dados da tabela 3.3, os resultados do item (a) são os mesmos do item (a) da tabela 3.2, assim, de acordo com o teste de Wilcoxon, rejeita-se a hipótese nula de igualdade entre as variáveis, confirmando a discordância entre elas.

Com os dados da mesma figura, agora do item (b), encontra-se o valor de Wilcoxon $W = 0$. Logo, a hipótese nula é rejeitada, mostrando que há diferenças entre as variáveis X e Y e por isso são discordantes, apesar de $\tau_{Kendall} = 1$ afirmar que existe associação perfeita entre elas. Logo, conclui-se que o teste de Wilcoxon é útil para a verificação de discordância (caso a hipótese de igualdade seja rejeitada) entre duas observações.

Note que ambos testes citados (correlação de Kendall e Wilcoxon) não são capazes de comprovar a concordância, quando utilizados separadamente. No entanto, a concordância pode ser verificada combinando esses dois testes. Um exemplo é apresentado abaixo, através dos dados (concordantes) da tabela 3.4.

Tabela 3.4: Valores de X e Y para combinação de Kendall e Wilcoxon

X	Y
1	1,1
2	2,1
3	3,1
4	4,1
5	5,1
6	6,1
7	7,1
8	8,1
9	9,1
10	10,1

Com base nos dados da tabela 3.4, quando as variáveis X e Y são analisadas, percebe-se que o teste de Kendall teve como resultado $\tau_{Kendall} = 1$, rejeitando a hipótese nula de que não existe associação entre elas. A concordância ainda não pode ser afirmada, já que o resultado também depende do teste de Wilcoxon. Este, por sua vez, com $W = 45$, conclui que a hipótese nula não deve ser rejeitada, ou seja, não há diferença entre as variáveis (entre os postos dos valores das variáveis). Assim, ambos os testes mostram que X e Y estão associados e compartilham postos similares, resultados que levam à conclusão que X e Y são concordantes entre si.

A tabela 3.5 apresenta os resultados para a combinação do coeficiente de correlação de Kendall e o teste de Wilcoxon dentro de um contexto de verificação da concordância entre duas medidas:

Tabela 3.5: Relação entre coeficiente de correlação de Kendall e teste de Wilcoxon

		Wilcoxon	
		Rejeita H_0	Não rejeita H_0
Kendall	Rejeita H_0	Discordância	Concordância
	Não rejeita H_0	Discordância	Discordância

Portanto, quando o teste da correlação de Kendall rejeita a hipótese nula (isto é, $\tau_{Kendall} > 0$) e o teste de Wilcoxon não a rejeita a igualdade de X e Y , a concordância entre as variáveis ordinais pode ser verificada.

Script do R

```
## DISCORDANTE
x<-seq(1,10)
y<-10+x
cor.test(x,y,method="kendall")
wilcox.test(x,y)
```

```
## CONCORDANTE
x<-seq(1,10)
y<-x+0.1
cor.test(x,y,method="kendall")
wilcox.test(x,y)
```

```
## DISCORDANTE
x<-seq(1,10)
y<-100-x
cor.test(x,y,method="kendall")
wilcox.test(x,y)
```

```
## DISCORDANTE
x<-c(1,2,3,4,5,6,7,8,9,10)
y<-c(10,1,9,2,8,3,7,4,6,5)
cor.test(x,y,method="kendall")
wilcox.test(x,y)
```

3.2.2 Coeficiente de Concordância de Kendall

O coeficiente de concordância de Kendall analisa a concordância entre variáveis ordinais. Caso $W_{Kendall} = 0$, pode-se afirmar que as variáveis em estudo são discordantes entre si.

Para ilustrar o cálculo e teste de $W_{Kendall}$, a tabela 3.6 apresenta conjuntos independentes de postos atribuídos por $k=3$ julgadores (X, Y, Z) referente a $N=6$ objetos (a a f).

Se existir uma concordância perfeita entre os postos, as somas de postos R_j , $j = 1, 2, \dots, k$ seriam a sequência $k, 2k, \dots, Nk$, para os $k = 6$ conjuntos de postos.

Para os dados da tabela 3.6, a média dos R_j é 10,5. Assim, por (2.10),

$$s = (8 - 10,5)^2 + (14 - 10,5)^2 + \dots + (8 - 10,5)^2 = 25,5$$

Tabela 3.6: Postos atribuídos à 6 objetos por 3 julgadores

Postos	Indivíduos					
	a	b	c	d	e	f
X	1	6	3	2	5	4
Y	1	5	6	4	2	3
Z	6	3	2	5	4	1
R_j	8	14	11	11	11	8

Fonte: Sidney Siegel [4]

Logo,

$$W_{Kendall} = \frac{25,5}{\frac{1}{12}(3)^2(6^3 - 6)} = 0,16$$

Calculando a estatística do teste (2.12),

$$Q_{Kendall}^2 = k(N - 1)W_{Kendall} = 2,43 \quad (3.3)$$

Assim, o nível descritivo do teste ($p - valor$) é dado por

$$p - valor = P(\chi_{(5)}^2 > 2,43) = 0,787 \quad (3.4)$$

O valor $W_{Kendall} = 0,16$ exprime o grau de concordância entre X , Y e Z . Como $p - valor = 0,787$, a hipótese de discordância não pode ser rejeitada. Assim, pode-se considerar que há discordância entre os três julgadores.

Script do R

```
install.packages("irr")
require(irr)
```

```
dados <- matrix(c(1,6,3,2,5,4,1,5,6,4,2,3,6,3,2,5,4,1), ncol = 3, nrow=6)
```

```
kendall(dados, correct = TRUE)
```

3.3 Variáveis Quantitativas

3.3.1 Teste T-Pareado e Coeficiente de Correlação de Pearson

O teste T-Pareado e o coeficiente de correlação de Pearson são utilizados para verificar, respectivamente, a diferença de médias e a associação entre duas variáveis quantitativas normais. De forma análoga ao coeficiente de correlação de Kendall e o teste de Wilcoxon, o uso de cada teste separadamente pode, no máximo, inferir sobre a discordância entre elas. Contudo, ao serem utilizados de forma conjunta, a concordância pode ser verificada.

O coeficiente de Pearson verifica se as variáveis estão ou não associadas entre si. Sozinho, nada pode concluir sobre a concordância, apesar do fato de que quando sua hipótese nula não é rejeitada, ou seja, quando conclui-se que as variáveis não estão associadas ($r_{Pearson} = 0$ ou $r_{Pearson} < 0$), a discordância entre elas é verificada.

Por exemplo, de acordo com a tabela 3.7, no item (a), ao se calcular o teste T-Pareado, verifica-se que a hipótese nula é rejeitada, ou seja, existem diferenças entre as duas variáveis. Porém, para o item (b), de acordo com o teste, não existe diferença entre elas, o que afirmaria que possuem a mesma média.

Tabela 3.7: Valores das variáveis X e Y para a correlação de Pearson

(a) T-Pareado rejeita H_0		(b) T-Pareado não rejeita H_0	
X	Y	X	Y
1	99	1	10
2	98	2	1
3	97	3	9
4	96	4	2
5	95	5	8
6	94	6	3
7	93	7	7
8	92	8	4
9	91	9	6
10	90	10	5

Analisando esses mesmos dados com o coeficiente de Pearson, $r_{Pearson} = -1$ e $r_{Pearson} = -0,151$ nos exemplos (a) e (b) respectivamente. Assim, ambas as correlações são negativas e seus p -valor $< \alpha$, considerando $\alpha = 0,05$, concluindo que as variáveis X e Y , de ambos os exemplos, são discordantes entre si, independentemente do resultado do T-Pareado.

Da mesma forma, quando o teste T-Pareado é realizado e sua hipótese de diferenças iguais entre as variáveis é rejeitada, apenas a discordância entre as variáveis quantitativas pode ser comprovada.

De acordo com o exemplo da tabela 3.8, o coeficiente de correlação de Pearson, para o item (a), assume que não existe associação entre as variáveis, enquanto que para o item (b), afirma que as variáveis apresentam associação. Agora, realizando o teste T-Pareado para os dois exemplos, tem-se que $T_{obs} = -65,730$ e $T_{obs} = -7,385$, respectivamente, o que mostra que independentemente dos valores de Pearson, quando o T-Pareado conclui que existem diferenças entre X e Y , pode-se afirmar que essas duas variáveis são discordantes entre si.

Tabela 3.8: Valores das variáveis X e Y para o teste T-Pareado

(a) Kendall não rejeita H_0		(b) Coeficiente de Pearson rejeita H_0	
X	Y	X	Y
1	99	1	11
2	98	2	12
3	97	3	13
4	96	4	14
5	95	5	15
6	94	6	16
7	93	7	17
8	92	8	18
9	91	9	19
10	90	10	20

Ao se combinar o coeficiente de correlação de Pearson com o teste T-Pareado, a concordância pode ser verificada. Considerando o exemplo da tabela 3.9, percebe-se que o $r_{Pearson} = 1$, rejeitando a hipótese de que não existe associação entre as variáveis X e Y e $T_{obs} = -0,074$, não rejeitando a hipótese de que não existe diferença entre elas.

Tabela 3.9: Valores de X e Y para combinação do teste T-Pareado com o coeficiente de Pearson

X	Y
1	1,1
2	2,1
3	3,1
4	4,1
5	5,1
6	6,1
7	7,1
8	8,1
9	9,1
10	10,1

Logo, a combinação de ambos os testes levam à conclusão de que as variáveis são concordâtes: as variáveis estão associadas positivamente e, por conseguinte, possuem a mesma média.

Essas informações foram resumidas na tabela 3.10,

Tabela 3.10: Relação entre teste T-Pareado e coeficiente de Pearson

		T-Pareado	
		Rejeita H_0	Não rejeita H_0
Pearson	Rejeita H_0	Discordância	Concordância
	Não rejeita H_0	Discordância	Discordância

em que,

para o coeficiente de Pearson,

$$H_0: r_{Pearson} \leq 0$$

$$H_a: r_{Pearson} > 0$$

e para o teste T-Pareado,

$$H_0: \mu_D = 0$$

$$H_a: \mu_D \neq 0$$

Portanto, quando o teste T-Pareado não rejeita a hipótese nula e Pearson a rejeita, a concordância entre as variáveis quantitativas pode ser verificada.

Script do R

```
## DISCORDANTE
```

```
x<-seq(1,10)
y<-10+x
cor.test(x,y,method="pearson")
t.test(x,y)
```

```
# CONCORDANTE
x<-seq(1,10)
y<-x+0.1
cor.test(x,y,method="pearson")
t.test(x,y)
```

```
## DISCORDANTE
x<-seq(1,10)
y<-100-x
cor.test(x,y,method="pearson")
t.test(x,y)
```

```
## DISCORDANTE
x<-c(1,2,3,4,5,6,7,8,9,10)
y<-c(10,1,9,2,8,3,7,4,6,5)
cor.test(x,y,method="pearson")
t.test(x,y)
```

3.3.2 Regressão Linear Simples

A análise de regressão linear simples tem como objetivo verificar a relação entre duas variáveis, uma explicativa e outra resposta. No entanto, para se inferir sobre a existência, ou não, de concordância entre elas, é necessário que se utilize um modelo sem intercepto ($\beta_0 = 0$). Caso essas duas variáveis concordem entre si, o valor do parâmetro $\hat{\beta}_1$ deve estar muito próximo de 1.

Considera-se uma amostra (tabela 3.11) cujas variáveis X e Y são quantitativas.

Aqui, X é a variável explicativa e Y é a variável resposta.

A concordância pode ser confirmada através do teste para verificar se o modelo de regressão linear possui $\beta_1 = 1$.

Calculando a estimativa do coeficiente angular da reta de regressão ($\hat{\beta}_1$),

Tabela 3.11: Valores de X e Y para a Regressão Linear

X	Y
23,1	10,5
31,8	18,2
30,4	16,3
39,5	23,1
52,5	24,9
30,5	14,1
12,4	8,8
31,5	14,9
27,6	16,1
32,8	16,7

Fonte: Weinsberg. S. [15]

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} = 0,5175$$

Assim, a estatística do teste pode ser encontrada da seguinte forma:

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{V(\hat{\beta}_1)}} = -23,7897$$

Calculando o do p-valor, tem-se:

$$p - \text{valor} = 2P(T_{(8)} > |-23,7897|) \approx 0$$

Considerando um nível de significância $\alpha = 0,05$, tem-se que $p - \text{valor} < \alpha$. Assim, conclui-se que X e Y não apresentam concordância entre si.

Script do R

```
x <- c(23.1,31.8,30.4,39.5,52.5,30.5,12.4,31.5,27.6,32.8)
y <- c(10.5,18.2,16.3,23.1,24.9,14.1,8.8,14.9,16.1,16.7)

n=length(x)

plot(x,y)
abline(lm(y~x-1))
```

```
beta1h <- sum(x*y)/sum(x^2)
varbeta1h <- ((sum((y-beta1h*x)^2))/(n-2))/(n*(sum(x^2)-n*(mean(x)^2)))
sdbeta1h <- sqrt(varbeta1h)

tobs <- (beta1h-1)/sdbeta1h

pvalor <- 2*pt(-abs(tobs),df=n-2)
qt(0.975,df=12)
```

Capítulo 4

Dois exemplos de aplicação dos testes de concordância

4.1 Aplicação 1 - Comparação de dois protótipos

Uma clínica é projetada para avaliar a possibilidade de compra de dois produtos, baseando-se nos seus protótipos. Assim, entrevistados avaliaram dois protótipos de produtos. Após a avaliação estar completa, cada um dos que analisaram o produto responderam se os comprariam. A clínica tem como interesse avaliar se as opiniões dos entrevistados concordam entre os dois protótipos, ou seja, se a possibilidade de compra pelos entrevistados é a mesma tanto para o protótipo A quanto para o B.

A tabela 4.1 abaixo mostra que 152 entrevistados concordaram que considerariam comprar ambos os produtos. Outros 155 concordaram em que não considerariam comprar nenhum deles [8].

Tabela 4.1: Tabela de contingência para a análise de protótipos

Consideram a compra do protótipo A	Consideram a compra do protótipo B		
	Sim	Não	Total
Sim	152	31	183
Não	61	155	216
Total	213	186	399

Fonte: Maritz Research

A tabela 4.1 é composta por variáveis nominais dicotômicas. Logo, o teste de concordância pode ser realizado por meio do Coeficiente de Concordância de Kappa.

De acordo com (2.3), o coeficiente de concordância é $\kappa = 0,5414$, que resulta em

$$p - \text{valor} = P(Z > Z_{\kappa}) = P(Z > 10,875) \approx 0$$

em que Z é uma variável com distribuição normal padrão e Z_{κ} é a estatística do teste dada por (2.4). Como $p - \text{valor} < \alpha = 0,05$, pode-se afirmar que a concordância

(apesar de ser apenas moderada) é significativa, isto é, clientes que comprariam (não comprariam) o protótipo A tendem, também, a comprar (não comprar) o protótipo B.

Script do R

```
prototipos <- matrix(c(152, 61, 31, 155),
nrow = 2,
dimnames = list("Comprariam o protótipo A" = c("Sim", "Não"),
"Comprariam o protótipo B" = c("Sim", "Não")))

install.packages(psych)
library(psych)

kappa <- wkappa(prototipos)
```

4.2 Aplicação 2 - Concordância em notas atribuídas por dois professores

Em uma universidade, ao final do semestre letivo é aplicada uma prova final, subjetiva, à 20 alunos, com o objetivo de analisar os conhecimentos adquiridos durante as aulas. Para isso, dois professores foram responsáveis pela correção de cada uma das provas. Pelo fato das respostas serem subjetivas, as notas podem diferir entre os dois professores. O objetivo aqui é verificar se as notas atribuídas pelos dois professores concordam ou discordam entre si.

Na tabela 4.2 estão as notas dos 20 alunos da disciplina. Como os dados estão representados por uma variável quantitativa, pode ser realizado: teste T-Pareado com o Coeficiente de Correlação de Pearson, a regressão linear e os gráficos de Bland-Altman.

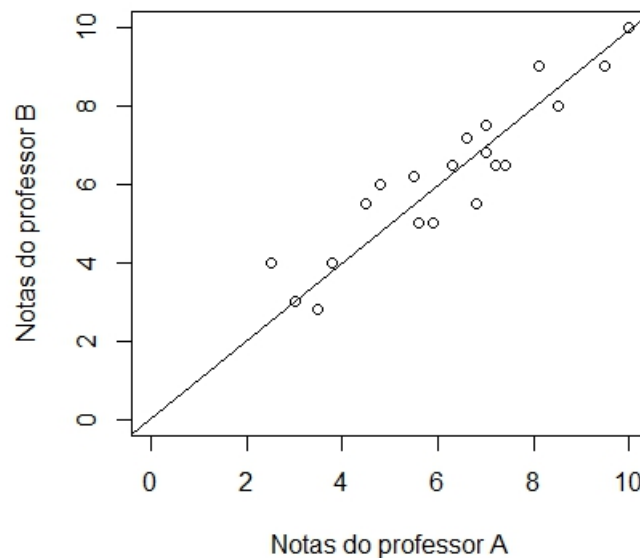
Tabela 4.2: Notas de 20 alunos dada por dois professores

Aluno	Notas		Aluno	Notas	
	Professor A	Professor B		Professor A	Professor B
1	8,5	8,0	11	7,4	6,5
2	3,5	2,8	12	5,6	5,0
3	7,2	6,5	13	6,3	6,5
4	5,5	6,2	14	3,0	3,0
5	9,5	9,0	15	8,1	9,0
6	7,1	7,5	16	3,8	4,0
7	4,8	6,0	17	6,8	5,5
8	6,6	7,2	18	10,0	10,0
9	2,5	4,0	19	4,5	5,5
10	7,0	6,8	20	5,9	5,0

Como apresentado em (2.19), $p\text{-valor} \approx 0$, quando se realiza o teste do coeficiente de Pearson e, por (2.21), $p\text{-valor} = 0,9688$ pelo teste T-Pareado. Como visto pela tabela (3.10), esse resultado implica na concordância das notas atribuídas pelos dois professores.

O mesmo pode ser verificado ao testar $H_0 : \beta_1 = 1$ em um modelo de regressão linear simples. Por (2.24), tem-se que $p\text{-valor} = 0,6664$, não rejeitando a concordância das notas atribuídas pelos dois professores e o mesmo pode ser verificado pelo gráfico de dispersão abaixo:

Figura 4.1: Gráfico de dispersão para as notas dos 20 alunos



Pelo que foi visto na seção 2.8, se a diferença entre as duas variáveis estiver dentro dos limites do intervalo, como mostra a Figura (4.2), significa que são concordantes entre si.

Assim, de forma quantitativa, as notas dadas à uma prova subjetiva pelos professores A e B concordam entre si.

As informações das notas dos alunos pode ser transformada em qualitativas ordinais quando apresentadas em postos. Essa transformação é útil quando as medidas quantitativas não seguem uma distribuição Normal. Nesta situação, os testes quantitativos (combinação dos testes t-Pareado e Pearson, regressão linear e gráficos de Bland-Altman) não são adequados para a verificação da concordância, sendo necessário a utilização da combinação dos testes de correlação de Kendall e Wilcoxon ou o coeficiente de concordância de Kendall. A Tabela (4.3) apresenta os postos das notas atribuídas pelos dois professores.

Figura 4.2: Gráfico de Bland-Altman para as notas dos 20 alunos

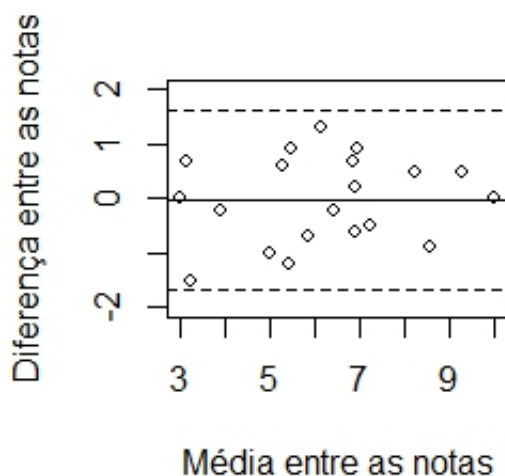


Tabela 4.3: Postos das notas de 20 alunos dada por dois professores

Aluno	Postos		Aluno	Postos	
	Professor A	Professor B		Professor A	Professor B
1	18	17	11	16	12
2	3	1	12	8	6
3	15	13	13	10	11
4	7	10	14	2	2
5	19	19	15	17	18
6	14	16	16	4	4
7	6	9	17	12	8
8	11	15	18	20	20
9	1	3	19	5	7
10	13	14	20	9	5,0

Para os dados da tabela (4.3), a concordância pode ser verificada por meio da combinação dos testes de Kendall e Wilcoxon e, de acordo com a seção (3.2.1) e com o cálculo dos p -valores (2.9) e (2.16), $p - \text{valor} \approx 0$ e $p - \text{valor} \approx 1$, respectivamente. De acordo com a tabela (3.5), esse resultado implica na concordância das notas atribuídas pelos dois professores.

De forma alternativa, a concordância pode ser verificada pelo coeficiente de concordância de Kendall (3.2.2). Para os dados da tabela (4.3), o p -valor do teste do coeficiente de concordância de Kendall (2.13) é dada por $p - \text{valor} = 0,0093$, indicando concordância das notas atribuídas pelos dois professores.

Considerando os dados representados da forma abaixo, de forma dicotômica (alunos aprovados são aqueles que obtiveram nota maior ou igual à 5), pela Tabela (4.4) o coeficiente de concordância de Kappa também pode ser encontrado.

Assim, considerando 2.5, $p - \text{valor} = 0,0049$ o que faz com que se rejeite a hipótese

Tabela 4.4: My caption

		Notas do professor B		
Notas do professor A		Aprovados	Reprovados	Total
	Aprovados	14	0	14
	Reprovados	2	4	6
	Total	16	4	20

nula de que não existe concordância entre as duas notas.

Neste exemplo, pôde-se concluir que, ao considerar um nível de significância de $\alpha = 0,05$, todos os testes realizados resultaram na mesma conclusão, de que as notas dos professores concordam entre si. Isto significa que, nesse situação, é necessário apenas um dos professores para avaliar as provas dos alunos.

A tabela abaixo apresenta um resumo dos resultados dos testes realizados para a verificação da concordância das notas atribuídas pelos dois professores.

Tabela 4.5: Resumo dos resultados dos testes realizados

Variável	Teste Utilizado	P-valor	Resultado
Qualitativa Nominal	Kappa	0,0049	Concordante
Qualitativa Ordinal	Correl. Kendall	<0,0001	Concordante
	Wilcoxon	1,0000	
	Conc. Kendall	0,0093	Concordante
Quantitativa	Pearson	<0,0001	Concordante
	T-Pareado	0,9688	
	Bland-Altman	-	Concordante

Script do R

```
p1 <- c(8.5,3.5,7.2,5.5,9.5,7,4.8,6.6,2.5,7,7.4,5.6,6.3,3,8.1,3.8,6.8,10,4.5,5.9)
p2 <- c(8,2.8,6.5,6.2,9,7.5,6,7.2,4,6.8,6.5,5,6.5,3,9,4,5.5,10,5.5,5)
```

```
### (T-Pareado + Pearson)
cor.test(p1,p2,method='pearson')
```

```

t.test(p1,p2)

### (Regressão)
n=length(p1)=length(p2)
beta1h <- sum(p1*p2)/sum(p1^2)
varbeta1h <- ((sum((p2-beta1h*p1)^2))/(n-2))/(n*(sum(p1^2)-n*(mean(p1)^2)))
sdbeta1h <- sqrt(varbeta1h)
tobs <- (beta1h-1)/sdbeta1h
pvalor <- 2*pt(-abs(tobs),df=n-2)

plot(p1,p2,xlim=c(0,10),ylim=c(0,10))
abline(lm(p1~p2-1))

### (Bland-Altman)
D <- p1-p2
media <- (p1+p2)/2
n=length(p1)
Sd <- sqrt(sum((D-mean(D))^2)/(n-1))
LS <- mean(D) + qt(0.975,19)*Sd
LI <- mean(D) - qt(0.975,19)*Sd
graf <- plot(media,D,xlab='Média entre as notas',ylab='Diferença entre as notas',
abline(h=mean(D))
abline(h=LS,lty=2)
abline(h=LI,lty=2)

### (Kendall + Wilcoxon)

A <- c(18,3,15,7,19,14,6,11,1,13,16,8,10,2,17,4,12,20,5,9)
B <- c(17,1,13,10,19,16,9,15,3,14,12,6,11,2,18,4,8,20,7,5)

cor.test(A,B,method="kendall")
wilcox.test(A,B)

### (Concordância Kendall)

notas <- matrix(cbind(A,B),nrow=20,ncol=2,
dimnames = list("Alunos" = c('1','2','3','4','5','6',
'7','8','9','10','11','12','13','14','15','16','17','18','19','20'),"Professores"

```

```

kendall(notas, correct = TRUE)#pvalor<alpha - rejeito. CONCORDÂNCIA!

### Kappa

dados <- matrix(c(14,2,0,4),nrow=2)

install.packages("psych")
require(psych)

k <- wkappa(dados)

pe <- ((16/20)*14/20)+((4/20)*6/20)
zk <- k$kappa/(sqrt(pe/(20*(1-pe))))

pvalor <- 1-pnorm(zk)

```

Capítulo 5

Conclusões

Tendo como base o trabalho realizado, pode-se considerar que todos os testes citados apresentam algum tipo de restrição, principalmente quanto ao nível de mensuração da variável analisada. Assim, é necessário discriminar as técnicas que pressupõem uma escala quantitativa (em geral, tem-se a suposição adicional de que as observações devam seguir uma distribuição Normal) com aquelas que podem ser utilizadas em uma escala qualitativa. Neste último caso, ainda é necessário subdividir as variáveis em dois critérios: ordinais e nominais.

Casos em que se deseja analisar a concordância de duas medidas quantitativas, Altman e Bland [16] descreveram exemplos mostrando que a comparação de médias, a correlação ou a regressão linear simples são procedimentos inadequados para a verificação da concordância e sugere que nesta situação uma análise gráfica pode ser realizada através do gráfico de Bland-Altman. Neste contexto, este trabalho propôs dois procedimentos alternativos para a verificação da concordância de duas medidas quantitativas quando os mesmos seguem uma distribuição Normal: 1) através de um teste de hipóteses do coeficiente angular de uma regressão linear simples sem intercepto e; 2) através de uma combinação do coeficiente de correlação de Pearson com o teste t-pareado.

Quando as variáveis em estudo seguem uma escala qualitativa ordinal, a concordância pode ser verificada através do coeficiente de concordância de Kendall ou, de forma alternativa, por meio de uma combinação do coeficiente de correlação de Kendall com o teste de Wilcoxon (análogo a combinação do coeficiente de correlação de Pearson com o teste t-parado no caso normal). Como todos esses testes de variáveis qualitativas ordinais são baseados em postos, os mesmos podem ser também aplicados em variáveis quantitativas. Desta forma, eles podem ser alternativas para os testes de variáveis quantitativas quando as mesmas não seguem a distribuição Normal.

Em último caso, quando as duas médias são qualitativas nominais, o teste de concordância de Kappa é o indicado para a verificação da concordância. O coeficiente de concordância de Kappa pode ser utilizado em todas as situações citadas acima, ou seja,

pode ser utilizado qualquer que seja a escala de medida da variável. Para a aplicação do coeficiente de concordância de Kappa nestas situações, basta categorizar (agrupar) os valores das (sejam quantitativas ou ordinais) em apenas duas categorias.

Como vistos nos scripts apresentados nas seções de ilustrações e aplicações dos testes de concordância, esses testes já estão disponíveis ou podem ser facilmente implementados no software R [3]. Essa facilidade de implementação pode incentivar o uso do R, que é um programa livre, principalmente por parte dos pesquisadores aplicados.

Referências Bibliográficas

- [1] CONOVER, W.J. **Practical Nonparametric Statistics**. 3 ed. New York: J. Wiley, 1999. 584 p. ISBN 0471160687.
- [2] NOETHER, GOTTFRIED E. **Introdução à Estatística: Uma abordagem não-paramétrica**. 2. ed. Rio de Janeiro: Guanabara Dois, 1983. 258 p.
- [3] R Core Team (2014). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3 – 900051 – 07 – 0, < [http : //www.R – project.org](http://www.R-project.org) >.
- [4] SIEGEL, S.; CASTELLAN, N.J. **Estatística Não-Paramétrica para as Ciências do Comportamento**. 2 ed. Porto Alegre: Artmed, 2006. 448 p. ISBN 9788536307299.
- [5] LANDIS, JR.; KOCH, G.G. **The measurement of observer agreement for categorical data**. Biometrics, 1977.
- [6] COHEN, J. A. **Coefficient of Agreement for Nominal Scales**. New York: Educacional Psychology, 1960.
- [7] VIEIRA, A.J.; GARRETT, J.M. **Understanding Interobserver Agreement: The Kappa Statistic**. Family Medicine, 2005.
- [8] Maritz Research - McNemar Test Example. Disponível em:
< [http : //www.maritzresearch.com/maritzstats/HelpFiles/..%5CHelpFiles%5CExamples_McNemarTest.htm](http://www.maritzresearch.com/maritzstats/HelpFiles/..%5CHelpFiles%5CExamples_McNemarTest.htm) > Acesso em 4 de fevereiro de 2015.
- [9] SALKIND, N.J. **Encyclopedia of Measurement and Statistics**. 1. ed. London: SAGE Publications, 2007. 508 p. ISBN 1 – 4129 – 1611 – 9.
- [10] Eberly College of Science. Disponível em:
< [https : //onlinecourses.science.psu.edu/stat509/node/159](https://onlinecourses.science.psu.edu/stat509/node/159) > Acesso em 19 de março de 2015.

- [11] VIALI, L. **Estatística Não-Paramétrica**. Série: Exatas.
- [12] Graham Hole Research Skills. Disponível em:
 $\langle \text{http} : // \text{www.sussex.ac.uk/Users/grahamh/RM1web/WilcoxonHandoout2011.pdf} \rangle$ Acesso em 14 de março de 2015.
- [13] FIGUEIREDO, D.; SILVA, J. **Desvendando os mistérios do coeficiente de correlação de Pearson**. Revista Política Hoje, Vol. 18, n. 1, 2009.
- [14] MOORE, David S., **The Basic Practice of Statistics**. New York, Freeman, 2007.
- [15] WEISBERG, S. **Applied Linear Regression**. 3. ed. Wiley Series, 2005. ISBN 0 – 471 – 66379 – 4.
- [16] ALTMAN, D.G.; BLAND, J.M. Source. **Measurement in Medicine: The Analysis of Method Comparison Studies**. Journal of the Royal Statistical Society, 1983. 307-317p. v. 32
- [17] ALTMAN, D.G.; BLAND, J.M. **Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement**. London: The Lancet, 1986.